DOCUMENT RESUME

ED 335 991 HE 024 791

AUTHOR Benson, Martin T., Ed.
TITLE A Statistical Analysis of the Predictions of Graduation Rates for College Student Athletes. NCAA Academic Performance Study. Report 91-02.
INSTITUTION National Collegiate Athletic Association, Overland Park, KS.
PUB DATE Jun 91
NOTE 34p.; For other reports in this series, see HE 024 789-790.
PUB TYPE Reports - Research/Technical (143)

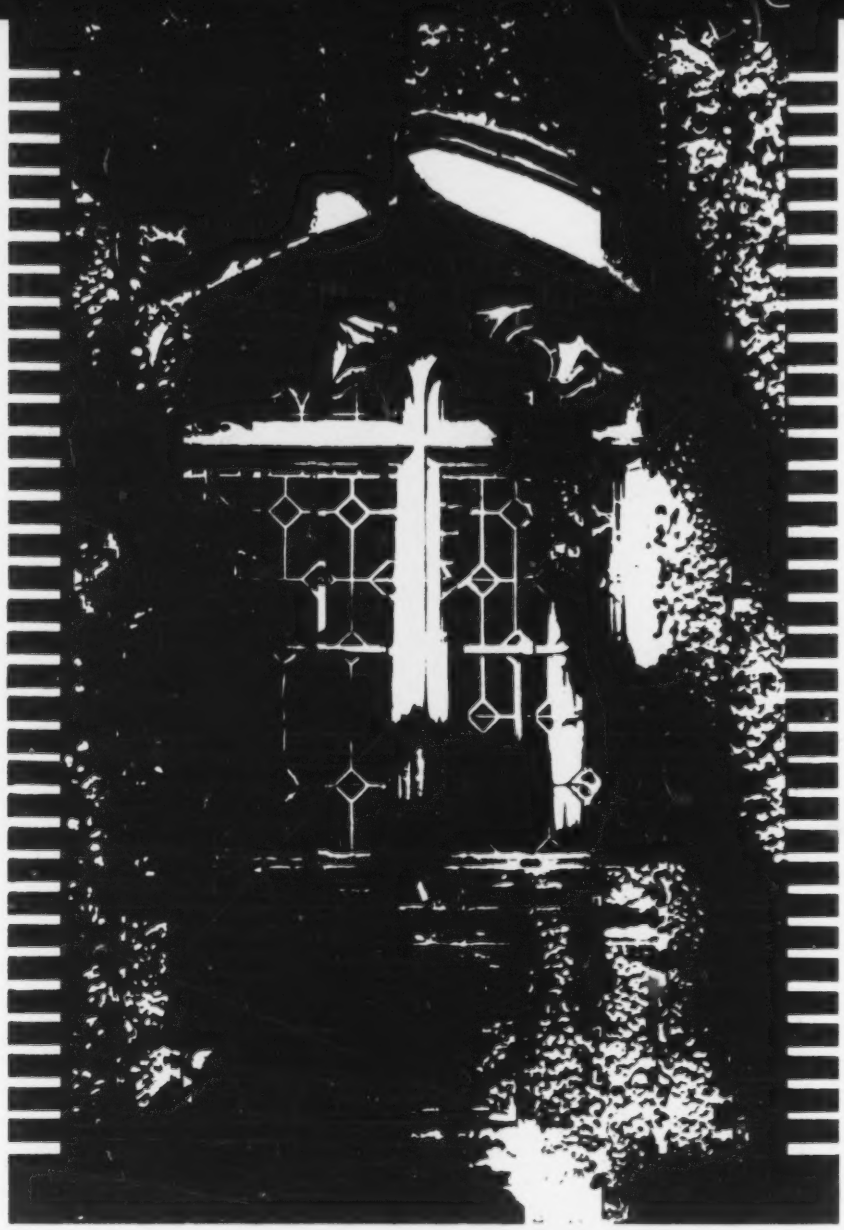
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Academic Achievement; *Athletes; Cohort Analysis; College Athletics; *College Graduates; *College Outcomes Assessment; Comparative Analysis; Grade Point Average; Higher Education; Minority Groups; *Predictor Variables; Standards; Statistical Analysis; Student Evaluation; Test Results
IDENTIFIERS National Collegiate Athletic Association; *Student Athletes

ABSTRACT

This report examines the predictors of college graduation for student athletes. Presented first are 5-year graduation results for the 1984 and 1985 cohorts of the National Collegiate Athletic Association (NCAA) data base. Questionnaires were sent to representatives at Division I schools in each of the 5 years. Data is analyzed from the responses of 82 different schools involving 4,845 student athletes. Table 1 summarizes the patterns of college outcomes and also gives breakdowns of graduation rates for white, black, male, and female student athletes. Table 2 provides a summary of descriptive statistics for 1984 and 1985 student athletes. The overall graduation rate as a probabilistic function of single predictor variables for 1984 and 1985 cohorts is provided in table 3. Figures for the linear combination of several of these variables are shown in table 4, and table 5 adds moderator variables and their interactions to the prediction equation. Table 6 is a summary of the accuracy and fit improvement of the high school grade point average and standardized test score variables broken down into separate demographic groups. Table 7 shows the improvement gained by adding knowledge of the institution that a student athlete chooses to attend, and the difference between the average student at the school and the student athlete in question. The final table provides a chronological summary of academic persistence prediction for 1984 and 1985 student athletes. Contains five references. (GLR)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED335991



NCAA Academic Performance Study

Report 91-02

A Statistical Analysis of the Predictions of Graduation Rates for College Student-Athletes

HE 024 791

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

NCAA

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

FOREWORD

This report is the third in a series that we shall be publishing to inform our member institutions and others about our ongoing study of student-athletes' academic performance under Bylaw 14.3.

The results presented are preliminary. This study was begun in 1985 and still has several years before completion. We do, however, anticipate some initial recommendations in 1991.

We welcome your comments and suggestions on this report.

RICHARD D. SCHULTZ
NCAA Executive Director
June 1991

THE NATIONAL COLLEGIATE ATHLETIC ASSOCIATION

6201 College Boulevard
Overland Park, Kansas 66211-2422
913/339-1906
July 1991

2

Edited By: Martin T. Benson, *Publications Editor.*

NCAA, NCAA seal, NCAA logo and NATIONAL COLLEGIATE ATHLETIC ASSOCIATION are registered marks of the Association and use in any manner is prohibited unless prior approval is obtained from the Association.

TABLE OF CONTENTS

Page

Introduction 6

Methods 6

Results 9

Discussion 11

References 13

TABLES

1 *Five-Year Graduation Rates for Various Groups of Student-Athletes from the 1984 and 1985 Cohorts* 14

2 *A Summary of Descriptive Statistics for 1984 and 1985 Student-Athletes* 15

3 *Overall Graduation Rate as a Probabilistic Function of Single Predictor Variables for 1984 and 1985 Cohorts* 17

4 *Overall Graduation Rate as a Probabilistic Function of Multiple Predictor Variables for 1984 and 1985 Cohorts (N=3,288)* 18

5 *Overall Graduation Rate as a Probabilistic Function of CORE GPA, TEST, and Moderator Variables for 1984 and 1985 Cohorts* 19

6 *A Summary of Overall Graduation Rate as a Probabilistic Function of Double Predictor Variables in Separate Group Analyses for 1984 and 1985 Cohorts* 20

7 *Adding SCHOOL Characteristics to the Overall Graduation-Rate Prediction for 1984-1985 Cohorts (N=3,022 individuals and G=82 schools)* 21

8 *A Chronological Summary of Academic Persistence Prediction for 1984 and 1985 Student-Athletes (N=3,288)* . . . 22

FIGURES

1 *The Relationship Between CGPA and TEST for Four 1984 and 1985 Student-Athlete Groups* 23

2 *Logit Prediction of Graduation Rate from CGPA Only (with 99 percent confidence boundaries)* 24

3 *Logit Prediction of Graduation Rate from TEST only (with 99 percent confidence boundaries)* 25

4 *Logit Prediction of Graduation Rate from CGPA and TEST (with 99 percent confidence boundaries)* 26

5 *The Bivariate Frequency Between Student Test Scores and the Average Test Scores of the Colleges They Attend* 27

APPENDIX TABLES

1 *Correlations of Selected Variables (N=3,288)* 29

2 *Logit Regression Models for Statistical Analysis* 30

INTRODUCTION

This report examines the predictors of college graduation for student-athletes. Presented first are five-year graduation results from the 1984 and 1985 cohorts of the NCAA data base. These results are relatively new—this determination of a five-year graduation did not occur until the fall of 1989 for the 1984 cohort, and the fall of 1990 for the 1985 cohort. Some details of our methodological approaches to the analysis of college persistence are also presented. All following analyses used logit regression models to predict graduation rates from high school grades, test scores, and other measures of practical and theoretical interest. Logit models were used to represent more accurately the prediction of an outcome that was a probability or rate (see Manski & Wise, 1983; also Hosmer and Lemeshow, 1989).

Data collected by the NCAA Academic Performance Study (NCAA APS) are illustrated in Figure [1]. These scatter-plots show the relationship between scores on high school grade-point averages in the core curriculum (CGPA) and precollege test scores (i.e. ACT or SAT; termed TEST). These scores describe measurements made on student-athletes during their last year or two of high school (in 1983, 1984 or 1985). Each point in the plot represents a single individual from one of four groups of student-athletes: white students who graduated from college in five years (by 1989 or 1990); white students who did not graduate; black students who graduated, and black students who did not graduate. These scatter-plots illustrate the contents of the NCAA APS data base to be analyzed.

This study's findings indicate that an equally weighted combination of CGPA and TEST provides the best prediction of graduation. The findings also indicate that the other variables (e.g. race, sex, school) increase the prediction's accuracy, but not significantly.

METHODS

Subjects

The NCAA Academic Performance Study (APS) data base covers the college careers of five cohorts of student-athletes numbering almost 12,000 students. These were student-athletes from the entering classes of 1984 to 1988 in Division I institutions that responded to the NCAA-APS questionnaire. These data are longitudinal; here we analyze data from the first two cohorts (1984 and 1985) of this study (for more details see NCAA Research Report No. 90-01).

The APS questionnaires were sent to representatives at a stratified random sample of G=57 Division I schools in each of the five years, 1984-1988. Every NCAA Division I institution was assigned to one of the five cohorts. The school representative was asked to fill out a form on each student-athlete in the assigned freshman class who had received athletically related aid in either the first or second year. In 1984, usable responses were obtained from N=44 of the 57 institutions. In 1985, usable responses from N=38 of a different set of 57 institutions were obtained. The

total response rate over both years was 71.9 percent. Participation patterns are addressed later in the study. From these responses, unique forms on N=4,845 student-athletes from G=82 different schools were obtained initially.

These individuals were reported on for each of five years using a similar form (see NCAA Research Report 90-01). For the purposes of the analyses to follow, individual records were cleaned. In sequence, eliminated were those persons: (1) with implausible data (N=114); (2) with incomplete longitudinal records (N=1,112); (3) who were not United States citizens, (N=204); (4) who did not indicate either white or black as their racial background (N=127), or (5) who were missing both high-school grade-point average (CGPA) and a standardized test score (TEST). The biggest data loss came from incomplete longitudinal records. More information on this may be available in the future, but for this study, data on the N=3,288 student-athletes from the entering classes of 1984 and 1985 with complete longitudinal records were used.

Table [1] summarizes the patterns of college outcomes. In the first column, a specific group is labeled. The first row lists the overall sample and the second column shows that this group was composed of N=3,288 persons with complete data for a full five-year span (this, in turn, was 100 percent of the total sample; column 3). Of these N=3,288 student-athletes, it was observed that N=1,504 graduated in the fourth or fifth year (column 4). This means the overall graduation rate is listed as 45.7 percent (1504/3288, column 5). It also was observed that some of the non-graduates left in good standing (as determined by the institution report). When the graduation base rate is adjusted to eliminate those who left in good standing, a new baseline of N=2,198 and an adjusted graduation rate of 68.4 (1504/2198, column 6) is obtained. Table [1] also gives breakdowns of graduation rates for white student-athletes (52.3 percent) and black student-athletes (26.6 percent), male student-athletes (42.3 percent) and female student-athletes (53.9 percent), and a variety of other classifications to be described in detail in the next section.

Variables

A variety of variables was obtained on each individual (for complete details on the APS survey variables, see NCAA Research Report No. 90-01). For simplicity, these variables were defined as either (1) precollege (2) in-college or (3) graduation variables. A listing of all variables used is presented in Table [2A]; univariate summary statistics are presented in Table [2B]. Correlations among these variables are presented in Appendix [A1].

One key variable in the analyses was the high-school core grade-point average (CGPA). The focus was on the CGPA in the 11 core courses required under NCAA Bylaw 14.3. The averages were obtained from actual records for about 80 percent of the individuals in the study, while 17 percent were imputed from available high-school grade-point-average data, and three percent were imputed from the available TEST score. A standardized z-score was computed on national averages for 1984 entering freshmen (see Table [2A] for details).

A similar strategy was used to deal with precollege test scores. Student-athletes in 1984 had taken either (1) the SAT, (2) the ACT, (3) both the SAT and the ACT, or (4) neither test. To deal with this problem, a variable labeled TEST as a z-score from either the SAT total score, the ACT composite score, or the aver-

age of both national z-scores was defined. This TEST score was imputed from CGPA for five percent of these students.

Class RANK was calculated as the percentage of the numerical rank of the individual divided by the size of the high-school graduating class. (Later, to aid communication, this variable was divided by five, so that each unit represented a 20-percent shift in class rank.) This variable was imputed from CGPA for about 18 percent of this sample. One other variable was used to define precollege scholastic performance: the AVERAGE score listed is a simple numerical average of the CGPA z-score and the TEST z-score. This variable is equivalent to a linear combination model in which both variables are used in standard-score form with equal weights.

The last few group variables were defined as "dummy codes" in Table [3A]. SEX MALE was coded 1 if the student is male and 0 otherwise (female). RACE BLACK was coded 1 if the student was black and 0 otherwise (white). SPORT REVENUE was coded 1 if the student's main college sport was men's basketball or football and 0 otherwise (all others). TRAVEL TEAM was coded 1 if the student athlete was on the traveling squad for at least one game during the freshman season. TUTOR USE was coded 1 if the student received tutoring during the freshman year. COHORT 1985 was coded 1 if the student entered in 1985 and 0 otherwise (1984). The final two group variables were OVERALL and ADJUSTED graduation rates. In all cases, these scores were coded for the individual student-athlete as 1 for graduate or 0 for nongraduate. The only difference between the two variables was that the adjusted rate did not include students who remained in school or who were in good standing when they left the institution. In this form, graduation is a binary variable that can be aggregated in various ways to produce a graduation rate or probability.

Statistical Analysis

The focus in most analyses is on the prediction of academic performance as indicated by persistence at the same institution. In this case, as in many others, persistence was defined as a binary outcome (i.e., either "yes" or "no"). It is well known that this binary measurement creates some problems for standard parametric measures such as linear regression. For example, with binary dependent variables, the assumption of homoscedasticity (i.e., homogeneity of error variance around the regression line) is necessarily failed, and the model parameters can be biased. Second, predictions outside the bounds of reasonable percentages are likely to be made. Third, the random-error scores should follow a binomial distribution rather than a normal distribution, so statistical characteristics of the model may be corrupted.

In order to deal with these problems, a logit regression model was used to predict persistence rates from CGPA and TEST scores (for details on logit or logistic regression, see Manski & Wise, 1983, or Hosmer & Lemeshow, 1990). This logit model for the odds ratio was the standard linear regression model with a binomial distribution for the error scores. This model was fitted to the binary data of Table [1] to obtain maximum likelihood using the available PC-SAS LOGISTIC computer program (SAS Institute, 1990). A statistical estimate of the overall goodness of fit and, hence, the relative size of the variance of error, also could be obtained.

These logit models were employed to predict persistence rates using a parameter for the intercept (B0), the direct effect of CGPA (B1), the direct effect of TEST (B2), and the direct effect of any linear interaction of CGPA and TEST (B3). Alternative models were compared for the significance of the parameters and the overall goodness-of-fit. The B-weights of the logit regression in terms of probability

or rates of change also were re-interpretted. This led to a particularly intuitive interpretation of the model parameters—the resulting weight was the "rate of change" in the graduation rate associated with a one-unit change in the independent variable. More complete details of the logit model are presented in Appendix Table [A2].

RESULTS

Table [3] displays the results of these logit regressions using precollege variables to predict the outcome of graduation from college. There are 10 variables listed, all of which were found to add a statistically significant amount to the prediction of graduation. The table's first column provides the intercept term. The second provides the logit weight, both in raw and standardized form. The third column provides the base rate for each variable. The fourth column gives the proportion of change for a one-unit increase in the independent variable. Thus, a one-unit jump in core grade-point average provided a 16.5 percent increase in the probability that a person would graduate. In terms of standardized units, a single standard deviation increase in the grade-point average amounted to a probability change of 11.5 percent.

Probably the most important column, in terms of judging accurate predictors, is the final column, change in goodness of fit. This column quantifies the amount of accuracy gained by having knowledge of a person's score on one of these variables. Thus, knowing a person's core grade-point average increased the accuracy of our predictions 11.1 percent. This column shows that the core grade-point average and test-score variables provided the greatest increase in our knowledge. Specifically, the combination of CGPA and TEST provided the best prediction of graduation, a 15.0-percent fit increase for overall graduation rate or 23.0 percent for adjusted graduation rate.

Unlike linear regression, the proportional rate of change was not linear for all unit changes in the independent variable. This feature is easier to see in a plot of the model, and Figure [2] gives such a plot for CGPA. To obtain an estimate of graduation rate from this model, a score on the X axis is chosen and moved up vertically to the heavy set prediction line; by drawing a horizontal line to the Y axis, the graduation rate that corresponds to this CGPA can be found. For example, at CGPA z-score=0, the expected graduation rate=52.8 percent (i.e., this is precisely the intercept). At CGPA z-score=–1, the expected graduation rate=32.1 percent (i.e., a one-unit change down). At CGPA z-score=–2 (i.e., where the CGPA=2.0), the expected graduation rate=20.0 percent. This last rate change still spans one unit but it is less than the rate change in the preceding interval. This last feature reflects the nonlinearity in the probability model.

There are a few other features of this kind of plot. The two faint lines listed around the regression line are the upper and lower 99-percent confidence boundaries of the predicted estimates. This regression line was created by using the standard error of estimate and the distance of the X-score from the X-score mean. Finally, several alternative scalings of the X-axis were listed for convenience of discussion. These axes allow a change of score in any unit to be converted into a change in the corresponding graduation rate. The "TEST only" model predictions are plotted in Figure

[3], and the same was done for the AVERAGE score in Figure [4].

Table [4] is similar in format to Table 3, except that it provides figures for the linear combination of several of these variables. This table shows that that the combination of CGPA and TEST did as well in terms of prediction as other available combinations. Lines 11 and 12 also show that a combination of CGPA and TEST with optimum weights was only slightly better than the same coupling with equal weights. Thus, from a practical standpoint, it was not efficient to weight these two elements differently. The average of test and core grade point average is the single score that is an optimal predictor.

Table [5] adds moderator variables and their interactions to the prediction equation. Five of these variables were statistically significant in the equation. None of these added a great deal to the overall fit of the model.

The models of Table [5] explore the differences among subgroups in more detail. The first model adds RACE information together with the CGPA and TEST scores. This model obtains significant coefficients for CGPA, and TEST, but not for the RACE, RACE BY CGPA or RACE BY TEST interactions, and the fit is improved to only 15.5 percent. This result can be interpreted to mean that, for all students athletes, the CGPA and TEST were useful predictors, but all other group differences were accounted for by these variables. The group differences and the two interaction coefficients would reflect any difference in the regression equation needed for the black student-athletes. Thus, there was no evidence here of the need for different equations; i.e., there was no "differential validity" between these racial groups on graduation rate. There were clear differences between the racial groups on graduation rate, but these were largely predictable from the differences in CGPA and TEST scores (see Hartigan & Wigdor, 1989). This is not to say that this evidence indicates no bias. The analyses employed here may not be not fully diagnostic of either bias or lack of bias.

Table [6] is a summary of the accuracy and fit improvement of the CGPA and TEST variables broken down into separate demographic groups. This is useful in that one can easily recognize the groups of student-athletes for which we were best able to predict future outcomes for the variables listed. The group differences defined above allowed the models to vary in the parameter estimates but not in the error variability. Specific BASELINE, CGPA, and CGPA and TEST models have been fitted to all main groupings of race, sex, and sports, as well as to each of the six separate groups. Table [6] gives an overview of these analyses, and few differences are observed.

Table [7] was added to assess the improvement gained by adding knowledge of the institution that a student-athlete chooses to attend and the difference between the average student at the school and the student athlete in question. This knowledge added about one percent more accuracy to the prediction. This addition of school level information may be useful in the actual functioning of the prediction.

The relationship between the TEST scores of STUDENTS and SCHOOLS is a frequently discussed issue, but not much data has been produced to define this relationship. Figure [5] gives a graphic display of the relationship between the STUDENT TEST scores on the SAT and ACT and the SCHOOL TEST average of the entering class. This relationship was reflected in a positive correlation (r=.54) between STUDENT TEST and SCHOOL TEST. There are two noticeable aspects to this figure. First, there is a large area of spread around the middle-level test scores, and all

student athletes seem to be going to all levels of schools in this range of score. Second, there is a smaller group of student athletes with the highest possible TEST scores, and these individuals seem to be going only to those schools with the highest possible TEST scores. This effect is probably smaller than expected because student athletes "self selected" to schools, or the opposite, and this yielded a positive distribution of the scores (as in Figure [5]). It remains to be seen whether this piece of information will be of value in later equations.

The final table summarizes the prediction of academic persistence from CGPA and TEST on a year by year basis. This table indicates that the test did not add much to the equation until the fourth or fifth year, but during the fifth year the addition was significant, and its contribution to the prediction was actually greater than that of the core grade point average.

DISCUSSION

These empirical results lead to the following summary conclusions:

1. The summary statistics for graduation rate (47.5 percent) and adjusted graduation rate (68.4 percent) are remarkably similar to other recent national statistics (e.g., Porter, 1990). The rates are also consistent for females (Overall=53.9 percent, Adjusted=79.4 percent) and for Blacks (Overall=26.6 percent, Adjusted=41.6 percent).

2. Graduation rate is moderately well predicted from both core grade point average and test score, and other variables. The rate of change predicted from TEST (13.4 percent) is larger than that of CGPA (11.5 percent), and together they represent up to a 15 percent improvement above a random chance base line. In general, the adjusted graduation rate shows the same pattern as the unadjusted, but is easier to predict (23 percent). An equally weighted average of core grade point average and test scores predicts as well as any other single or multiple variable examined here (with a rate of change of 13.9 percent).

3. There are clear precollege score differences between black and white student athletes. The Blacks are lower on both core grade point average and test scores, and their graduation rate is also lower. However, the prediction equations for both groups show similar slopes and error variance, so there is little or no evidence for differential validity of prediction between racial groups (see Hartigan & Wigdor, 1989).

4. There are group differences in the graduation rate between the sexes. The females graduate at a higher rate than the males, but there is no evidence for differential validity between the sexes.

5. Other precollege and in-college variables used in this study do not add much to these predictions. High school rank is similar to core grade point average in prediction, but rank does not significantly add to the prediction. There are small differences between sports teams, cohorts, students who made the travel team as freshmen, and students who received tutoring during

their freshman year, but these variables have no major effect on the main results.

6. Average level of test scores of schools is a potentially useful predictive factor, but individual test scores are more important. This may be evidence for self-selection of students to schools, or vice versa.

7. The graduation-rate predictions increase in accuracy over time. This change is most notable between the fourth and fifth year. This may be due to the fact that many students drop out after being eligible for three or four years.

8. This kind of study of fifth-year graduation rate is rare in educational literature. We have cross-validated all results over the two cohorts, and these results are stable.

9. The predictions we produced are moderate. Variables such as core grade-point average and test scores can help us cut down on the uncertainty about the effects of college-level preparation but they are far from perfect.

10. The equally weighted combination of core grade-point average and test score provides the single best prediction of graduation.

REFERENCES

Hosmer, D. W., & Lemeshow, S. (1989). *Applied logistic regression.* New York: Wiley.

Manski, C. F., & Wise, D. A. (1983). *College choice in America.* Cambridge, MA: Harvard University Press.

Porter, O. F. (1990). *Undergraduate completion and persistence at four-year colleges and universities.* Washington, D.C.: National Institute of Independent Colleges and Universities.

Tinto, V. (1987). *Leaving college.* Chicago: University of Chicago Press.

Hartigan, J. A. & Wigdor, A. K. (1989). *Fairness in employment testing: Validity generalization, minority issues, and the general aptitude test battery.* Washington, D.C.: The National Research Council and National Academy of Sciences Press.

TABLE 1

FIVE-YEAR GRADUATION RATES FOR VARIOUS GROUPS OF STUDENT-ATHLETES FROM THE 1984 AND 1985 COHORTS

Group Breakdown	Total Student Athletes		Graduated from School by Fifth Year	Overall Graduation Rate	Adjusted Graduation Rate
	N	%	N	(N=3,288)	(N=2,198)
OVERALL	3,288	100.0%	1,504	45.7%	68.4%
RACIAL GROUP					
Whites	2,453	74.6%	1,282	52.3%	77.0%
Blacks	835	25.4%	222	26.6%	41.6%
SEX					
Males	2,314	70.3%	979	42.3%	63.7%
Females	974	29.6%	525	53.9%	79.4%
COHORT					
1984 Cohort	1,780	54.1%	901	50.6%	74.1%
1985 Cohort	1,508	45.9%	603	40.0%	61.4%
FRESHMAN TRAVELING TEAM					
Did Travel	2,295	69.7%	1,095	47.7%	72.0%
Did Not Travel	993	30.2%	409	41.2%	60.3%
FRESHMAN YEAR TUTORING					
Did Receive Tutor	1,019	31.0%	345	33.9%	51.0%
Did Not Receive Tutor	2,269	69.0%	1,159	51.1%	76.2%
SEPARATE SPORTS					
Male Revenue Sports	1,296	39.6%	543	41.9%	62.5%
Male Other Sports	1,018	31.1%	436	42.8%	65.3%
Female Other Sports	956	29.2%	515	53.9%	79.5%
SEPARATE GROUPS					
White Male Revenue	751	22.8%	412	54.9%	78.5%
White Male Other Sports	895	27.2%	407	45.5%	69.3%
White Female All Sports	807	24.5%	463	57.4%	83.9%
Black Male Revenue	563	17.1%	141	25.0%	39.5%
Black Male Other Sports	123	3.7%	29	23.6%	35.8%
Black Female All Sports	149	4.6%	52	34.9%	54.2%

Notes: All percentages based on sample size listed. Adjusted graduation rate is based on total sample minus any dropouts who left in good standing or continuing students who remain in good standing (good standing determined by the school).

TABLE 2

A SUMMARY OF DESCRIPTIVE STATISTICS FOR 1984 AND 1985 STUDENT-ATHLETES

2a—A Summary List of Variables and Coding

PRECOLLEGE VARIABLES	
1. CGPA:	Standardized high school cumulative grade point average in the eleven core courses; ZCGPA=(HSGPAC-3.05) / .52; statistical constants from Klitgaard (1986) 1984 national sample
2. Precollege TEST:	Standardized precollege achievement test score based on SAT and ACT; ZSAT=(SAT-906) / 210; ZACT=(ACT-19.6) / 5.8; if both tests are available, ZTEST=(ZSAT+ZACT) / 2; Statistical constants from Klitgaard (1986) 1984 national sample
3. Class RANK:	Student's high school class ranking in proportion form (rank divided by the graduation class size).
4. AVERAGE Z Score:	Composite ability scores based on equally weighted CGPA and STUDENT TEST; (ZCGPA+ZTEST) / 2;
5. SEX MALE:	A binary variable representing male or female; coded as female=0 or male=1.
6. RACE WHITE:	A binary variable representing racial group; coded as black students=0 or white students=1.
IN COLLEGE VARIABLES	
7. SPORT REVENUE:	A binary variable representing main sport of the student-athlete; coded as 0 for most sports, but 1 for revenue sports (i.e., basketball=1, football=1).
8. TRAVEL TEAM:	A binary variable representing traveling experiences in freshman year; coded as did not make traveling team=0, or did make traveling team=1.
9. TUTOR USE:	A binary variable representing the use of extra tutoring in the freshman year; coded as did not use tutoring=0, or did use tutoring=1.
10. COHORT 1985:	A binary variable representing different cohort years; Coded as 1984 cohort=0, or 1985 cohort=1.
GRADUATION VARIABLES	
11. OVERALL Rate:	Individual graduation from participating colleges; coded as not graduated by fifth year=0 or graduated by fifth year=1.
12. ADJUSTED Rate:	Individual graduation adjusted for dropouts who left in good standing; (dropout not in good standing=0, or graduated=1, or dropout in good standing=missing).

2b—Univariate Summary Statistics for Selected Variables for All 1984 and 1985 Student-Athletes (N=3,288)

Variable Label	Raw Scale Mean	(St. Dev)	Rescaled Mean	Percent Missing
PRECOLLEGE VARIABLES				
1. CGPA	-.3649	(1.233)	CGPA=2.84	17.9%
2. Student TEST	-.1449	(1.083)	SAT=864, ACT=19.0	5.7%
3. Class RANK	.3434	(.229)	34.3% in class	18.1%
4. Z Score AVERAGE	-.2549	(1.053)	not applicable	23.6%
5. SEX MALE (vs Female)	.7038	(.457)	70.4% males	0.0%
6. RACE WHITE (vs Black)	.7461	(.435)	74.6% Whites	0.0%
IN-COLLEGE VARIABLES				
7. SPORT REVENUE (Main)	.3996	(.490)	40.0% revenues	0.0%
8. TRAVEL TEAM (Frosh)	.6970	(.458)	69.7% traveled	0.5%
9. TUTOR USE (Frosh)	.3100	(.479)	31.0% used tutors	13.3%
10. COHORT 1985	.4586	(.498)	45.9% of 1985	0.0%
GRADUATION VARIABLES				
11. OVERALL Grad Rate	.4574	(.499)	45.7% graduated	0.0%
12. ADJUSTED Grad Rate	.6840	(.489)	68.4% graduated	33.2%

Note: Percent missing based on N=3,288. Missing scores were imputed by simple linear regression. Adjusted rate statistics based on N=2,198.

TABLE 3

OVERALL GRADUATION RATE AS A PROBABILISTIC FUNCTION OF SINGLE PREDICTOR VARIABLES FOR 1984 AND 1985 COHORTS

Model Fitted	Logit Coefficients: Intercept Coefficient	Logit Coefficients: Weight Coefficient (Standardized)	Rates-of-Change: Base Rate	Rates-of-Change: Proportion of Change (Standardized)	Fit Index: Overall (N=3,288)	Fit Index: Adjusted (N=2,198)
0. Baseline	-.171**	.000 (.000)	45.7	0 (0.0)	0.0%	0.0%
1. CGPA	.054	.692** (.470)	51.4	16.5** (11.5)	11.1%	18.6%
2. Student TEST	-.077*	.918** (.548)	48.1*	21.8** (13.4)	13.9%	19.3%
3. AVERAGE Z Score	.042*	.980** (.575)	51.1*	22.4** (13.9)	15.0%	23.0%
4. Class RANK (20%)	.978**	-.692** (-.438)	72.7**	-15.6** (-10.8)	9.5%	14.7%
5. RACE BLACK	.091*	-1.106** (-.266)	53.9*	-25.7** (-6.6)	3.8%	7.4%
6. SEX MALE	.156*	-.467** (-.117)	53.9*	-11.6** (-2.9)	0.8%	1.8%
7. SPORT REVENUE	-.073	-.246** (-.067)	48.2	-6.1** (-1.7)	0.3%	1.0%
8. TRAVEL TEAM	-.356**	.265** (.067)	41.2**	6.5** (1.7)	0.3%	0.6%
9. TUTOR USE	.043	-.713** (-.182)	67.9	-23.4** (-4.5)	1.9%	3.6%
10. COHORT 1985	.025	-.431** (-.118)	50.6	-10.6** (-2.9)	0.8%	0.6%

Notes: All models were analyzed based on the combined sample of 1984 and 1985 presented. Logit coefficients are calculated using SAS 6.04 PROC LOGISITIC. Rates-of-change are calculated from standard exponential formulas (see Table [4]). Rates-of-change represent the change in probability of graduation from baseline for a one-unit change in the predictor variable. Standardized rates-of-change are calculated from standardized logit values. These hazard rates represent the change in probability of graduation for a one standard deviation change in the predictor variable. Fit index defines improvement in fit over the baseline. Asterisk designates statistical significance (i.e., * = $p < .05$, and ** = $p < .01$).

TABLE 4

OVERALL GRADUATION RATE AS A PROBABILISTIC FUNCTION OF MULTIPLE PREDICTOR VARIABLES FOR 1984 AND 1985 COHORTS (N=3,288)

Model	Rates of Change				Fit Index	
	Base Rate	CGPA	TEST	RANK	Overall	Adjusted
Additional Variables Model						
11. CGPA & TEST	50.3	8.4**	16.2**	—	15.4%	23.0%
		(5.7)	(10.2)	—		
12. Equal CGPA & TEST	51.0	22.5**	22.5**	—	15.0%	23.0
	(13.9)	(13.9)	—			
13. CGPA & RANK	60.7**	11.0**	—	-6.5**	11.5%	19.1%
		(8.3)	—	(-4.2)		
14. TEST & RANK	62.1**	—	15.0**	-8.6**	15.6%	22.3%
		—	(10.6)	(-5.6)		
15. CGPA, TEST & RANK	58.4**	4.2**	14.7**	-5.6**	15.8%	23.4%
		(3.0)	(9.8)	(-4.7)		

Notes: All rates of change calculated from logit coefficients (not presented). Asterisk designates statistical significance (i.e., *=$p < .05$, and **=$p < .01$).

TABLE 5

OVERALL GRADUATION RATE AS A PROBABILISTIC FUNCTION OF CORE GPA, TEST, AND MODERATOR VARIABLES FOR 1984 AND 1985 COHORTS

Model	Rates of Change (and Standardized)						Fit Index	
	Base Rate	CGPA	TEST	Group Variable	CGPA by Group	TEST by Group	Overall	Adjusted
16. Adding Race - Black (Black, Black×CGPA, Black×Test)	49.9	7.6** (5.2)	16.7** (10.2)	4.0 (1.0)	-4.7 (-1.7)	-1.1 (-0.4)	15.5%	23.2%
17. Adding Sex - Male (Male, Male×CGPA, Male×Test)	54.8	6.7** (4.7)	14.9** (9.5)	-6.8** (-1.7)	1.3 (0.7)	1.6 (0.8)	15.7%	23.5%
18. Adding Sport - Revenue (Revenue, Rev×CGPA, Rev×Test)	48.3	8.9** (6.1)	16.5** (10.0)	5.6* (1.5)	-.4 (-0.5)	0.5 (0.2)	15.6%	23.1%
19. Adding Travel - Team (Travel, Travel×CGPA, Travel×Test)	46.2	6.6** (4.6)	16.3** (9.8)	5.6* (1.4)	2.2 (1.3)	0.3 (0.2)	15.6%	23.2%
20. Adding Cohort 1985 (Cohort, Cohort×CGPA, Cohort×Test)	55.2	9.6** (6.8)	14.0** (8.9)	-10.9** (-3.0)	-3.5 (-0.8)	3.8 (1.5)	16.1%	23.4%
21. Adding Tutor (Tutor, Tutor×CGPA, Tutor×Test)	52.2	7.2** (4.9)	16.8** (10.5)	-7.8** (-2.0)	4.1 (1.6)	-6.0* (-2.1)	15.8%	23.9%
22. Adding Nonlinear ($CGPA^2$, $TEST^2$ and CGPA×Test)	51.1	7.7** (5.3)	16.6** (10.3)	-1.3 (-1.4)	-0.6 (-0.6)	2.1 (1.6)	15.5%	23.1%

Notes: Asterisk designates statistical significance (i.e., *=$p < .05$, and **=$p < .01$).

TABLE 6

A SUMMARY OF OVERALL GRADUATION RATE AS A PROBABILISTIC FUNCTION OF DOUBLE PREDICTOR VARIABLES IN SEPARATE GROUP ANALYSES FOR 1984 AND 1985 COHORTS

Group		Rates of Change			Fit Index	
	N	Base Rate	CGPA	TEST	Overall	Adjusted
RACE						
G1. White	2,453	49.9	11.8**	15.2**	12.7%	15.6%
G2. Black	800	53.9	7.6**	16.7**	12.0%	17.5%
SEX						
G3. Female	974	54.8**	6.7**	14.9**	10.6%	15.9%
G4. Male	2,314	48.1	8.1**	17.2**	16.9%	24.3%
SPORTS						
G5. Revenue	1,314	53.9*	8.9**	16.3**	17.4%	24.4%
G6. Other	1,974	48.3	8.9**	16.5**	14.0%	20.8%
TRAVEL						
G7. Traveling Team	2,295	51.2	8.9**	16.1**	16.1%	22.6%
G8. Did Not Travel	993	46.3*	6.9**	16.4**	13.6%	23.0%
TUTOR USE						
G9. Was Not Tutored	2,269	52.2	7.3**	16.8**	14.8%	22.2%
G10. Received Tutoring	1,019	44.5**	11.4**	11.8**	12.9%	18.6%
COHORT						
G11. Cohort 1984	1,780	55.2**	9.6**	14.0**	16.4%	23.9%
G12. Cohort 1985	1,508	44.3**	6.5**	18.7**	14.2%	21.7%
SIX GROUPS						
G13. White-Male Revenue	751	55.2*	6.9**	14.4**	11.1%	16.0%
G14. White-Male Other	895	39.5**	6.9**	22.3**	16.5%	22.9%
G15. White-Female	807	54.5*	5.9**	16.4**	9.5%	13.2%
G16. Black-Male Revenue	563	54.2	11.3**	15.7**	12.6%	16.2%
G17. Black-Male Other	123	51.7	6.2	25.5**	19.4%	22.8%
G18. Black-Female	149	51.0	14.0**	8.0	8.3%	8.9%

Notes: Asterisk designates statistical significance (i.e., *=$p < .05$, and **=$p < .01$).

TABLE 7

ADDING SCHOOL CHARACTERISTICS TO OVERALL GRADUATION RATE PREDICTION FOR 1984-1985 COHORTS (N=3,022 INDIVIDUALS AND G=82 SCHOOLS)

Model	Rates of Change					Improved Fit	
	Base Rate	CGPA	TEST	School	Other	Overall	Adjusted
S0. CGPA and Student TEST	51.3	8.3** (5.7)	15.6** (9.7)	— —	— —	15.0%	22.6%
S1. CGPA and School TEST	39.7**	12.7** (8.6)	— —	23.4** (9.2)	— —	15.7%	22.4%
S2. CGPA and Difference TEST	53.9**	15.1** (10.7)	— —	— —	3.0* (1.5)	10.9%	19.3%
S3. CGPA, School TEST, and Difference TEST	42.0**	7.6** (5.2)	— —	28.5** (11.4)	11.3** (5.8)	17.4%	24.2%
S4. CGPA, School TEST, and Student TEST	42.0**	7.6** (5.2)	11.3** (6.8)	18.2** (7.1)	— —	17.4%	24.2%
S5. CGPA, School TEST, and Instit Grad	14.2**	7.8** (8.9)	— —	-0.6 (-0.5)	2.5** (11.3)	18.7%	26.2%
S6. CGPA, School TEST, and Public vs Private	38.8**	13.0** (8.8)	— —	19.1** (7.5)	16.9** (3.8)	16.6%	23.7%
S7. CGPA, School TEST, and Division	45.5*	12.8** (8.6)	— —	23.7** (9.5)	-10.2** (-2.9)	16.3%	23.9%
S8. CGPA, School TEST, Student TEST, and Inst Grad	16.2**	6.5** (5.8)	5.3** (6.2)	-2.4 (-1.9)	2.6** (10.7)	20.1%	27.4%

Notes: All hazard rates calculated from logit coefficients (not presented). Asterisk designates statistical significance (i.e., *=$p < .05$, and **=$p < .01$).

TABLE 8

A CHRONOLOGICAL SUMMARY OF ACADEMIC PERSISTENCE PREDICTION FOR 1984 AND 1985 STUDENT-ATHLETES (N=3,288)

		Rates of Change		Fit Index	
Year	Base Rate	CGPA (Standardized)	TEST	Overall	Adjusted
P1. First Year Persistence (Proportion Persisting=80.0%)	82.6**	2.3** (2.9)	3.7** (4.2)	3.3%	9.2%
P2. Second Year Persistence (Proportion Persisting=64.7%)	68.3**	4.5** (3.7)	5.7** (4.2)	4.4%	9.9%
P3. Third Year Persistence (Proportion Persisting=63.1%)	66.9**	4.4** (3.5)	7.2** (5.2)	5.4%	12.8%
P4. Fourth Year Persistence (Proportion Persisting 55.6%)	60.3**	6.1** (4.4)	10.0** (6.6)	8.5%	15.5%
P5. Fifth Year Persistence (Proportion Persisting=50.1%)	55.2**	7.7** (5.4)	13.9** (8.8)	13.1%	21.6%

Notes: Asterisk designates statistical significance (i.e., *=p< .05, and **=p< .01).

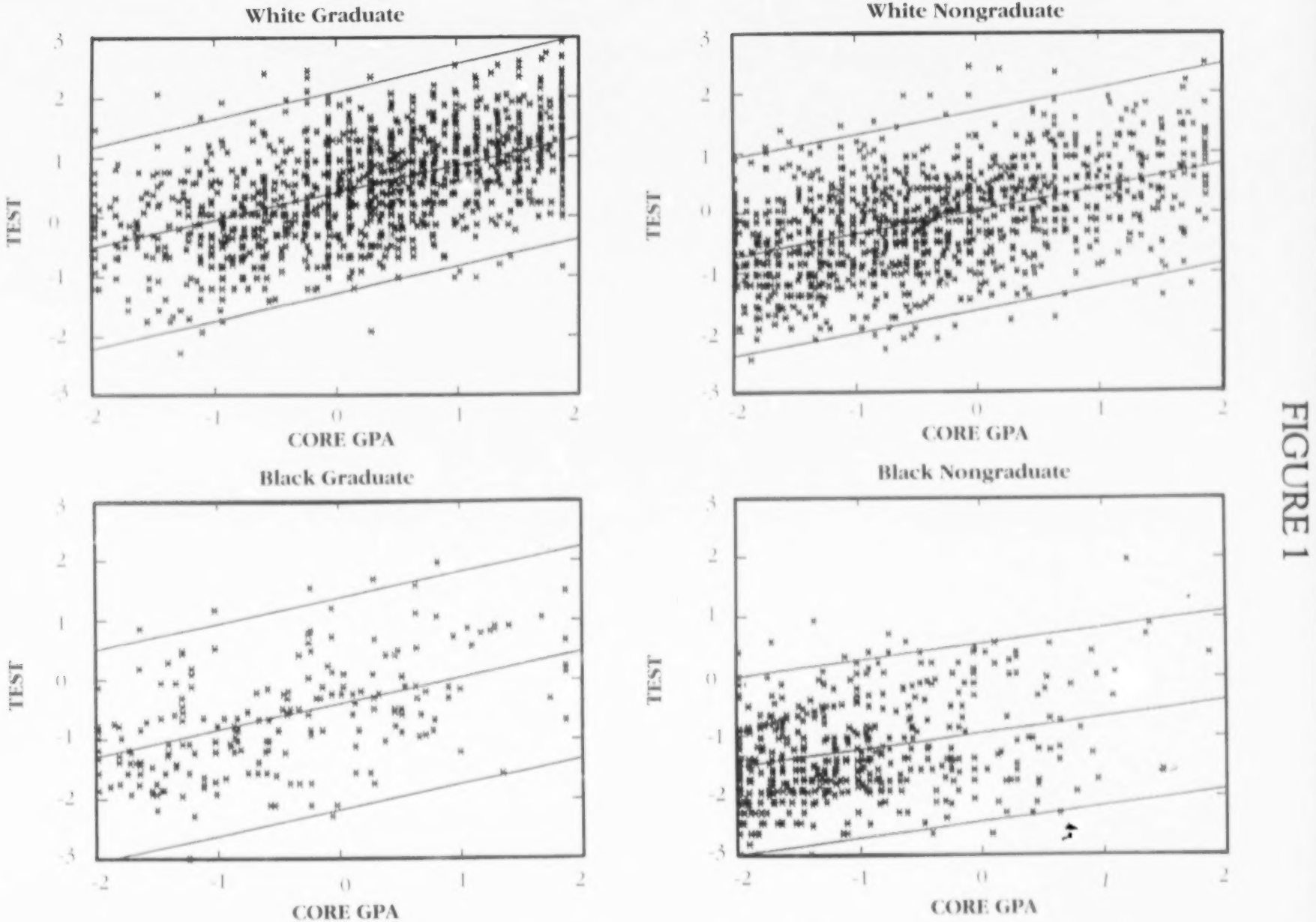
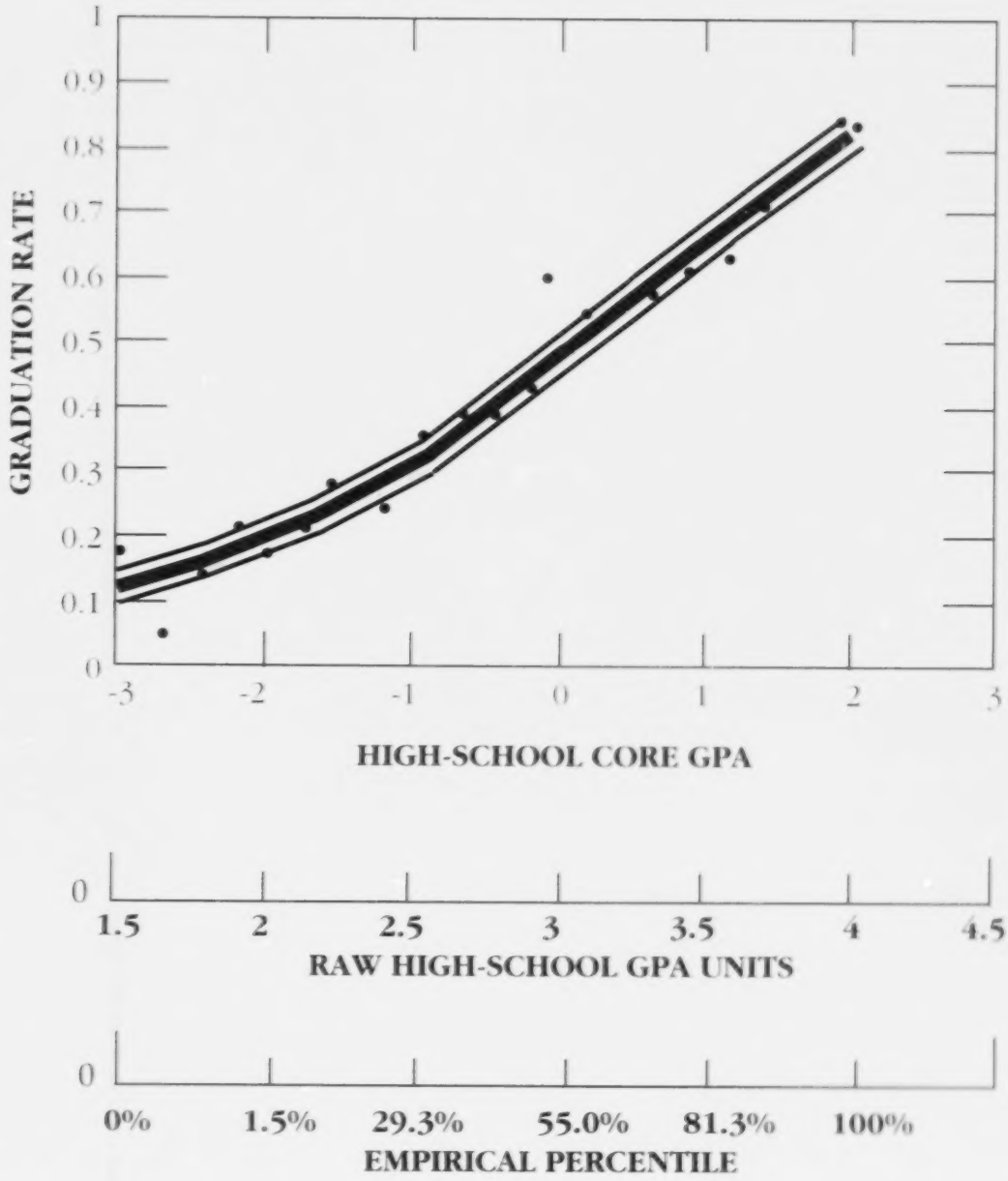


FIGURE 1

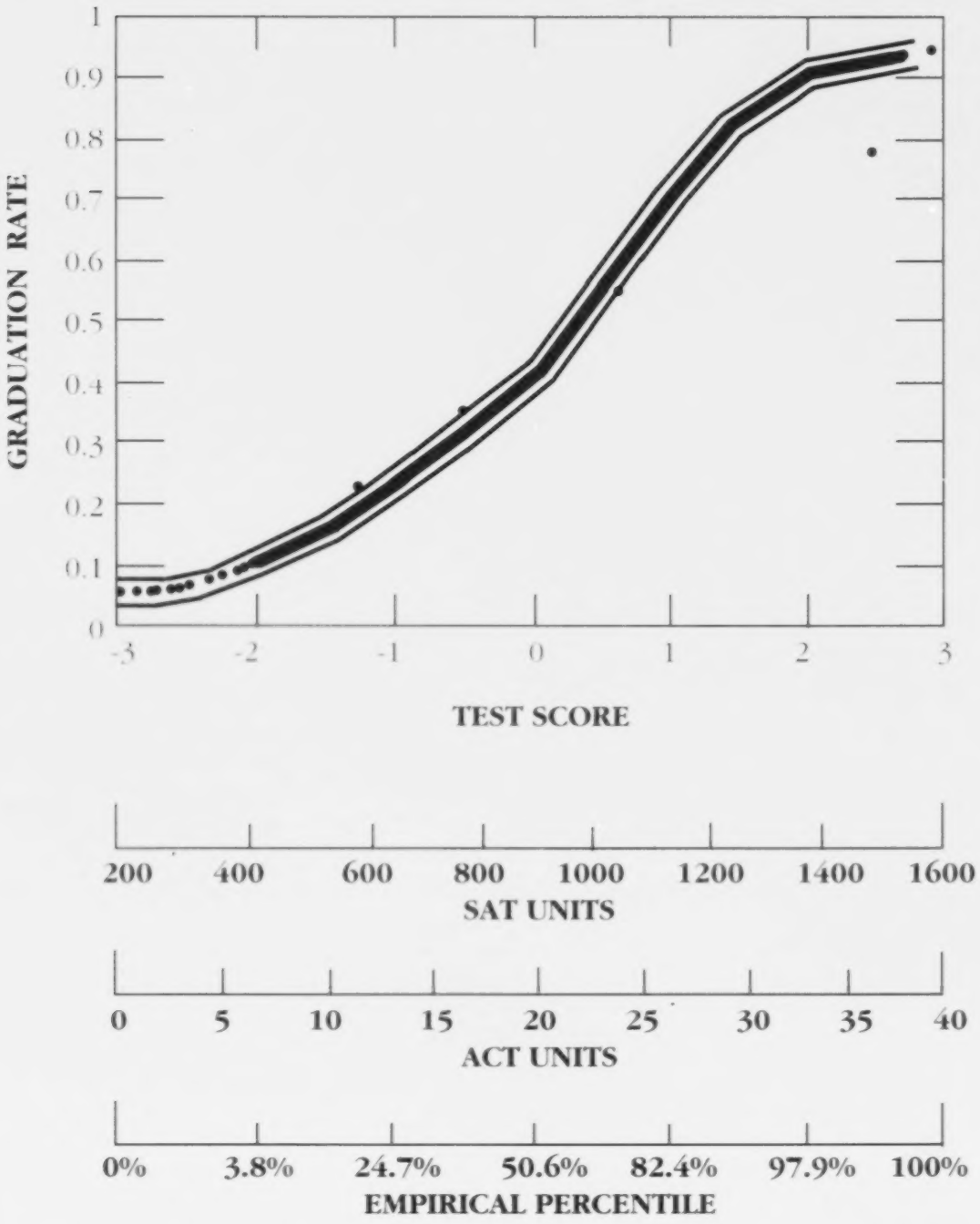
The relationship between CGPA and TEST for four 1984 and 1985 student-athlete groups.

FIGURE 2



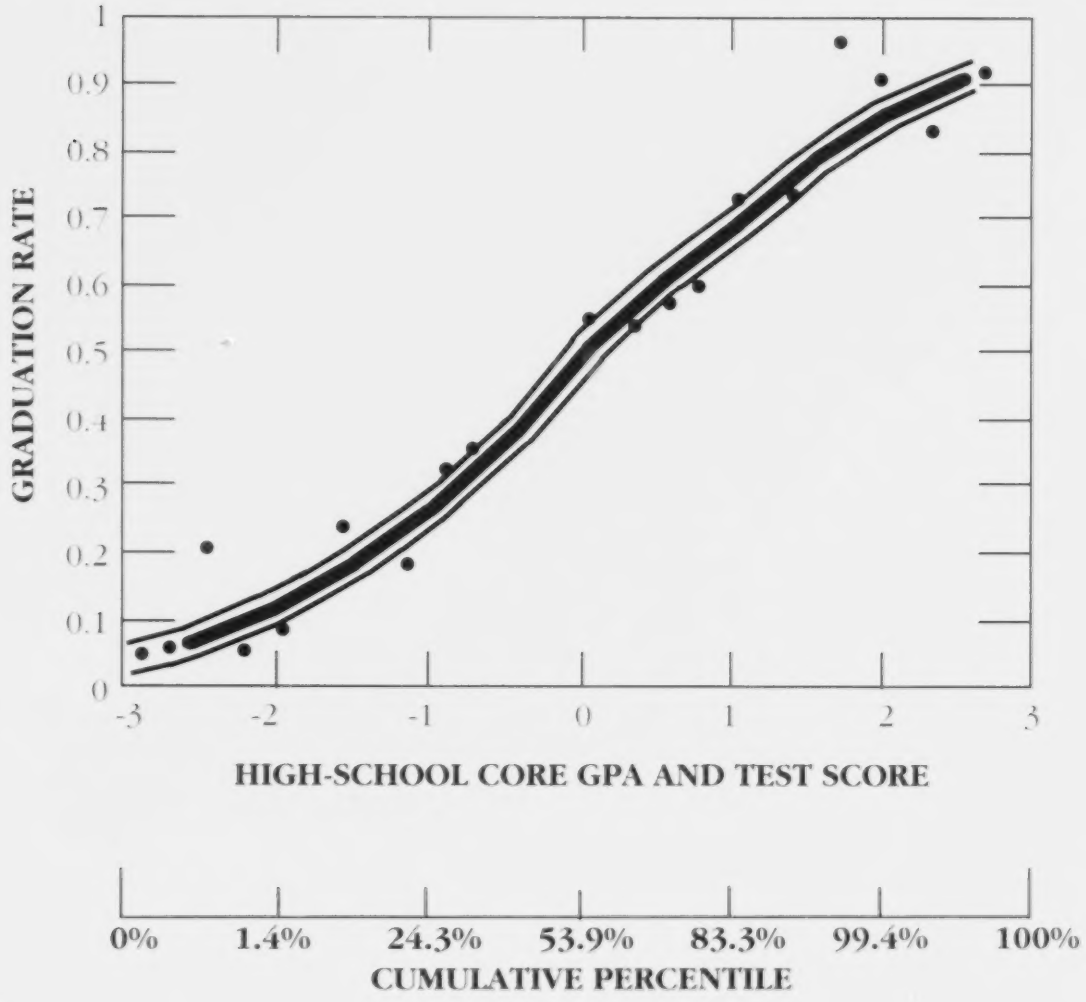
Logit prediction of graduation rate from CGPA only (with 99 percent confidence boundaries).

FIGURE 3



Logit prediction of graduation rate from TEST only (with 99 percent confidence boundaries).

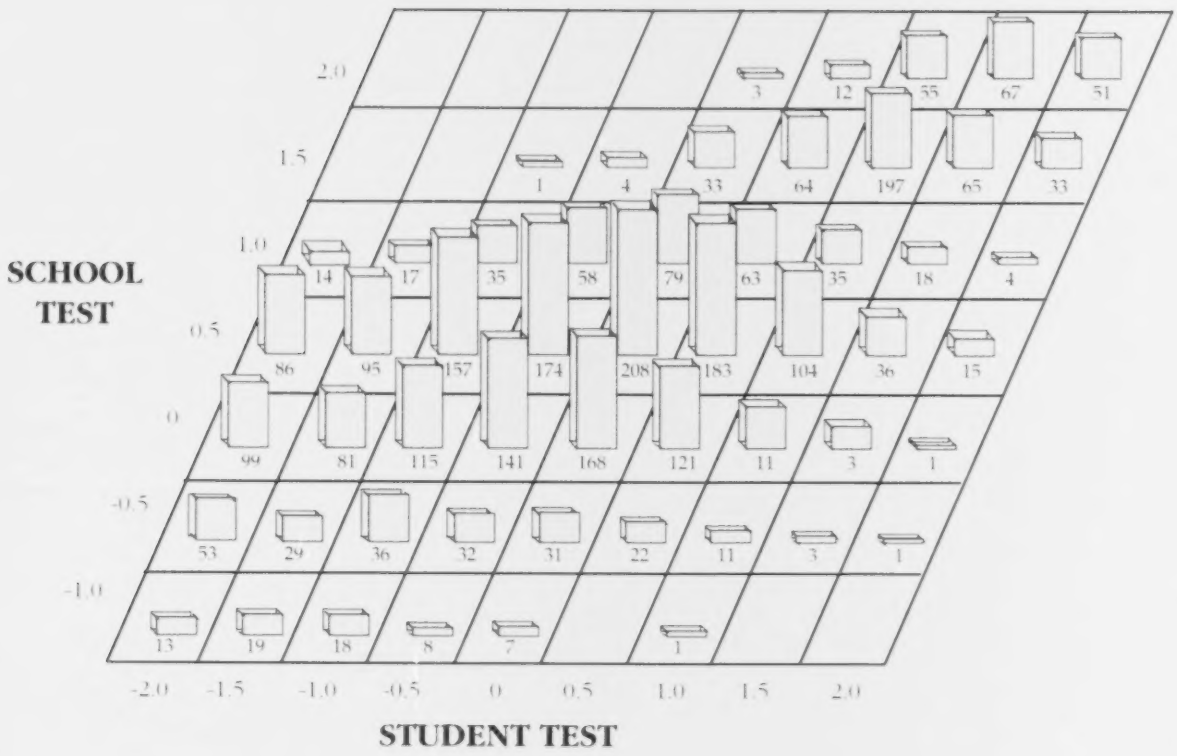
FIGURE 4



26

Logit prediction of graduation rate from CGPA and TEST (with 99 percent confidence boundaries).

FIGURE 5



The bivariate frequency between student test scores and the average test scores of the colleges they attend.

APPENDIX

APPENDIX TABLE 1

CORRELATIONS OF SELECTED VARIABLES (N=3,288)

Variables	CGPA	TEST	RANK	AVER	SEX	RACE	SPORT	TRAVEL	TUTOR	COHORT
1. CGPA	1.000									
2. TEST	.652	1.000								
3. RANK	-.803	-.553	1.000							
4. AVERAGE	.921	.896	-.755	1.000						
5. SEX	-.205	-.068	.207	-.155	1.000					
6. RACE	.377	.545	-.261	.501	-.140	1.000				
7. SPORT	-.208	-.222	.172	-.236	.505	-.327	1.000			
8. TRAVEL	-.077	-.046	.070	-.069	.303	-.032	.409	1.000		
9. TUTOR	.177	.284	-.175	.249	-.039	.224	-.193	-.074	1.000	
10. COHORT	-.065	-.045	.032	-.061	-.025	.010	-.066	.067	.067	1.000

Variables	CGPA	TEST	RANK	AVER	SEX	RACE	SPORT	TRAVEL	TUTOR	COHORT
11. OVERALL	.379	.421	-.351	.438	-.106	.224	-.060	-.056	.174	-.106
12. ADJUSTED	.475	.484	-.433	.522	-.154	.317	-.115	-.083	.233	-.092

Notes: Percent missing based on N=3,288. Missing scores were imputed by simple linear regression. Adjusted rate statistics based on N=2,198. All coefficients where abs(R)>.045 are significantly different than zero at the $p<.01$ test level.

APPENDIX TABLE 2

LOGIT REGRESSION MODELS FOR STATISTICAL ANALYSIS
(for more details, see Fox, 1984, or Hosmer & Lemeshow, 1990)

---

(1) The Dependent Variable:

Logit(n) = LOG ODDS(n) = LOG [G(n) / 1-G(n)] ,
where G(n) is graduation status for student n.
G(n) = 1 = graduated in five years from same college, or
G(n) = 0 = not graduated in five years from same college, and
Logit(n) = the dependent variable G(i) transformed into the natural logarithm (LOG) of the odds ratio (ODDS) of graduation.

(2) Several Linear Logit Regression Models:

[2a] Logit(n) = B0 + E(n), and
[2b] Logit(n) = B0 + B1 * CORE-GPA(n) + E(n), and
[2c] Logit(n) = B0 + B1 * CORE-GPA(n) + B2 * TEST(n) + E(n), etc.,
where B0, B1, and B2 = logit coefficients, and
E(n) = misfit of model to observed log odds.

(3) Probability or Rate-of-Change Transformation

[3a] Odds(n) = EXP[Logit(n)], so
[3b] Expected G(n) = [Expected Odds(n) / (1 + Expected Odds(n))] .
[3c] Base Rate = EXP(B0) / [1+ EXP(B0)]
[3d] Rate-of-Change = [EXP(B0+B1) / (1+ EXP(B0+B1))] - BASE RATE
where EXP( ) = the exponential or antilog function, so
Rate-of-Change = the change in the probability of the outcome for a one-unit change in the independent predictor variable away from the average score.

[4] Likelihood Ratio Fit Index (Pseudo Explained Variance)

$$LIP = \frac{[\ LR(B0) - LR(B0 + B1)\ ]}{LR(B0)} * 100$$

where LIP Likelihood Improvement as a Percentage;
LR(B0) = the likelihood ratio based on the intercept only, and
LR(B0 + B1) = the likelihood ratio based on the intercept plus an independent variable.

NCAA 8067-6/91